

이름	박태민	글쓰기 수업 분반	000200D-12
학번	20233216	글쓰기 담당 교수	박필현 교수님

2023학년도 1학기

글쓰기

기말과제

AI 그림의 숨겨진 문제점

-학습 과정에서의 인간과 AI의 차이점-

과목명	글쓰기(공학인증)
담당교수	박필현 교수님
학수번호	000200D-12
이름	박태민
학과	건축학부
학번	20233216
연락처	010-7553-9544
제출일	2023. 06. 11.

AI 그림의 숨겨진 문제점

-학습 과정에서의 인간과 AI의 차이점-

박태민(건축학부)

-목차-

- I. 서론
 - II. AI가 그림을 그리기 위한 학습 과정에서 발생하는 저작권 문제
 - III. AI와 인간의 학습 과정의 특징
 - 1. AI 학습 과정의 특징
 - 2. 인간 학습 과정의 특징
 - IV. AI와 인간의 학습 과정의 차이
 - V. 결론
- 참고문헌

I. 서론

최근 AI가 매우 빠른 속도로 좋은 질의 그림을 만들어내며 화제가 되고 있다. 예로부터 그림은 작가의 특정한 감정과 심상, 그리고 나타내고자 하는 바를 표현하는 의사소통의 방식과 예술 활동 등 다양한 형태로 많은 사랑과 관심을 가졌다. 오랜 시간 동안 그림이라는 예술, 창작 활동은 인간의 영역으로만 생각했었지만, 기술이 발달하며 AI가 그 경계를 침범함에 따라 여러 문제가 일어나고 있다. AI 그림에 관한 문제로는 AI의 그림에 저작권을 부여해야 하는가, AI가 일러스트레이터의 자리를 완전히 대체할 때까지 보고만 있을 것인가 등의 다양한 문제가 있다. 이 보고서에서는 그중 AI가 이런 그림을 그릴 수 있을 때까지 학습하는 과정에서 발생하는 저작권 문제에 관심을 가지고 살펴보려 한다. AI가 그림을 그리기 위해 웹상의 그림 데이터를 무단으로 가져와 학습하는 과정에서 세계적으로 많은 위법 사례가 있고, AI 학습 과정과 이에 대한 저작권 문제를 다룬 다양한 논문과 연구에서도 현재 AI 학습 과정과 관련된 저작권법의 문제를 이야기하며 이에 대한 논의와 연구를 요구하는 등 AI 그림 학습 과정의 저작권 문제는 점점 대두되고 있다. 그런데 여기서 인간의 사례를 생각해보면 의문이 생긴다. AI가 무단으로 여러 그림을 모아서 학습하는 것과, 인간이 평상시에 그림 실력을 늘리기 위해 웹상에 떠돌아다니는 아무 그림들을 보고 그리는 것에는 어떤 차이가 있길래 AI의 학습 과정에만 저작권 문제가 대두되는지이다. AI의 학습 과정에 대한 문제를 제기하는 논문과 연구는 있지만 인간과 AI의 학습 과정을 비교하는 연구는 찾아볼 수 없었다. 따라서 이 보고서에서는 AI와 인간의 학습 과정의 차이를 탐구하고 왜 AI의 학습 과정에서만 문제가 나타나는지 그 차이를 통해 답을 제시하려 한다. 이를 위해 2장에서는 현재 AI가 그림을 그리기 위해 학습하는 과정에서 발생하는 저작권 문제에 대해 알아보고, 3장에서 AI와 인간의 학습 과정의 특징을 탐구한 뒤 4장에서 그 차이를 찾아 이를 근거로 결론을 내려 한다.

II. AI가 그림을 그리기 위한 학습 과정에서 발생하는 저작권 문제

AI가 그림을 그리기 위해 학습하는 과정에서 데이터를 모으는 크롤링이란 행위에서 저작권 문제가 대두되고 있다. 웹사이트의 운영 의사에 반하거나 규정되어 있는 법이나 규칙을 어기는 불법적 크롤링에서 저작권법상 데이터제작자의 권리를 침해할 가능성이 있다. 그러나 자율성이 강한 AI의 경우 빅데이터나 저작물을 무단으로 수집하였을 때 그로 인해 피해를 입은 제작자의 권리를 보호하는 별도의 구제 수단이 현행법상에는 없기에 세계적으로 이와 관련된 논쟁이 계속 일어나고 있다. 이는 빠르게 변해가는 과학기술의 발전에 법의 변화가 맞추어 따라가지 못하고 있는 현상으로, 개별적인 재판 결과에 권리보호를 맡기거나, 이와 관련된 법을 새로 규정하는 예시들이 늘어나고 있다.

이와 관련된 대표적인 사례로는 먼저 2023년 1월과 2월 2번에 걸쳐 '게티이미지'가 'Stability AI'란 스타트업을 상대로 소송을 제기한 예시가 있다. 게티이미지는 Stability AI가 1,200만장 이상의 사진을 무단으로 복제하며 자신의 서비스와 경쟁하고 있다고 주장하며 영국의 법원과 미국 델라웨어주의 연방법원에 소송을 제기하였다. 그들은 피고가 웹상에서 예술가의 동의를 받지 않고 500억개의 이미지로 AI를 교육시켜 예술가의 저작권을 직접, 간접적으로 침해했다고 주장하였다. 그리고 이 분쟁에서 저작물이 복제, 전송된 것은 사실이므로 복제, 전송권의 침해와 AI 결과물이 원 저작물과 유사한 것을 바탕으로 2차 저작물 작성권이나 동일성 유지권과 같은 침해문제가 인정된다.

다음으로는 NovelAI라는 이미지 생성모델의 예시이다. 앞서 말한 Stability AI가 생성한 NovelAI의 하위 항목인 이미지 제너레이터는 사용자가 입력한 키워드나 문장을 통해 만화풍의 캐릭터를 AI가 생성하거나, 사용자가 그린 그림을 더 높은 품질로 수정하는 것이 가능한 이미지 생성 AI이다. NovelAI의 뛰어난 성능 때문에 2022년 10월 출시 직후 화제가 되었으나 이미지 학습을 'Danbooru'란 일본의 서브컬처 관련 이미지 사이트에서 무단으로 했다는 것이 드러나 저작권 문제 지적을 받고 있다. 또한 NovelAI는 유료 구독을 통해 제공되는 서비스라 무단으로 타인의 창작물을 이용해 학습시킨 AI의 기능을 유료로 판매한다는 점에서 논란이 계속 커지고 있다.

마지막 예시로는 일본의 '미믹'이란 서비스 논란이다. 미믹은 특정 창작자의 화풍을 학습해 그 화풍으로 새로운 캐릭터를 만들 수 있는 서비스이다. 최소 15장의 일러스트를 학습시키면 AI가 2시간 만에 화풍을 학습해 다양한 일러스트를 만들 수 있다. 원래 미믹은 창작자들이 참고 자료로 사용하거나 자신의 화풍으로 새로운 그림을 빠르게 만들어 팬들에게 선물하는 등의 용도로 사용하도록 만들어진 서비스였지만 창작자들이 해당 서비스가 남의 화풍으로 그림을 찍어내는 등의 우려가 있다며 반발하였고, 논란이 커지자 미믹은 공개된 지 하루만에 사과와 함께 서비스를 종료하였다.

미믹 논란 이후 일본에서는 창작자가 자신의 창작물을 AI 학습에 사용하지 말라고 요청하는 문구가 늘었다고 한다. 하지만 일본 내에서 2018년에 개정된 저작권법에는 AI 연구에서 학습과 빅데이터 분석에 쓰이는 데이터는 저작권과 관계없이 자유롭게 이용할 수 있어 미믹의 서비스에는 법률적인 문제가 없다고 한다. 일본뿐만 아니라 다른 나라에서도 AI가 창작한 결과물에 대한 논의와 법 개정은 활발하게 이루어지고 있지만 AI 학습 과정에 대한 저작권 문제는 과학기술 연구의 학습에 따른 문제로 분류되어 정확한 법적 재제가 없는 맹점이 존재한다.

Ⅲ. AI와 인간의 학습 과정의 특징

1. AI 학습 과정의 특징

AI가 그림을 그리기 위해 학습하는 과정에서만의 특징이 아니라, AI 학습에서 통용되는 전반적인 학습 과정의 특징에 대해 알아보려고 한다. AI의 학습 과정은 크게 5단계로 분류할 수 있다.

먼저 AI가 목표로 하는 학습에 맞는 데이터를 수집하는 과정이다. 이 과정에서 AI는 목표로 하는 바에 부합하는 최대한 많은 양의 데이터를 모두 수집한다. 목표로 하는 학습이 그림을 그리기 위한 학습이라면 최대한의 그림 데이터를 모두 수집하는 것이 이 과정에 해당한다. 두 번째로는 이렇게 수집한 빅데이터 속에서 목적에 맞는 양질의 데이터만을 걸러내는 것이다. AI 학습 연구에서 기술이 점점 발달함에 따라 데이터의 양보다는 질이 중시되고 있고, 때문에 무분별하게 많은 데이터를 분석하는 것보다는 그 목적에 맞는 소수의 양질의 데이터를 분석하는 것이 원하는 결과를 얻기에 더 적합하다고 여겨진다. 따라서 빅데이터 중 핵심 요건을 충족하는 데이터를 뽑아내고, 이 목적에 맞는 양질의 데이터들을 데이터 세트라 칭한다. 위에서 예시로 들었던 'NovelAI'의 경우 만화풍의 그림을 그려내는 것이 목표이기에 이 학습 과정에서는 수많은 그림 데이터 중 서양화, 팝아트 등의 데이터를 모두 제외하고 일본 만화풍의 그림만 뽑아낸다. 세 번째는 이 데이터 세트를 수천 번이 넘는 복제, 모방, 제복제를 통해 분석하는 것이다. 네 번째 단계는 이렇게 반복 학습을 통해 데이터 세트 속의 통계적인 규칙과 패턴, 특성을 찾아내는 것이다. 수많은 데이터 사이에서 일치하는 공통점을 찾아내는 것이 이 과정의 목표로, NovelAI를 예로 들자면 일본 만화풍의 그림은 대체적으로 눈동자를 어떻게 그리는지와 같은 특성을 찾아내는 것이다. 마지막으로 기존의 데이터를 분석해 찾은 패턴과 특성, 규칙을 통해 새롭게 들어올 데이터, 또는 입력값에 대한 예측값을 구하는 것이 AI 학습 과정의 최종적인 목표이다. NovelAI에 사용자가 원하는 입력어를 넣었을 때 이 입력어에 맞는 만화풍의 그림을 만들어 낼 수 있도록 5단계의 과정을 통해 AI를 학습시키는 것이다.

2. 인간 학습 과정의 특징

인간의 학습 과정의 시작은 외부에서 오는 자극을 눈, 귀, 촉각과 같은 감각기관으로 받아들이는 것이다. 받아들인 자극을 감각기관에서 신경세포를 통해 뇌로 보내 자극에 대한 지각을 하고, 이에 대한 판단과 분석을 내린다. 이때 이 자극은 우연적이거나 비의도적인 자극일 수도 있고, 목표에 따라 의도적으로 받아들이는 자극일 수도 있다. 이렇게 자극을 받아들여 뇌를 통해 판단, 분석을 내리는 과정은 AI가 데이터를 수집하여 분석하는 과정과 비슷하다고 볼 수 있다. 이렇게 내린 판단과 분석은 자극에 대한 반응을 내리는 데 쓰이고, 동시에 우리의 기억 속에 저장된다. 이 기억 속에 저장되는 판단과 분석이 후에 새로운 자극이 들어올 때 현재 상황에 대한 반응을 내리는 데 활용되는 인간의 학습 과정의 핵심이다. 간단한 예시로, 가시가 달린 공이 날라올 때 그것을 받았다가 아프다는 감각을 느껴 몸을 보호하기 위해 손을 바로 때는 반응을 했다면, 그 다음에 똑같은 상황이 일어났을 때는 전의 기억을 살려 손으로 받지 말고 피해야겠다는 실행계획을 세우고 피하는 반응을 보이는 것이다. 그림 실력을 늘리기 위해 그림을 잘 그리는 방법과 같은 정보를 찾아보고, 그림을 따라 그리며 기억 속에 저장된 분석과 경험 또한 후에 다른 그림을 그리는 상황이 찾아왔을 때 활용된다.

기억으로 저장된 자극과 정보에 대한 분석은 인간의 학습 과정의 핵심적인 요소이지만, 인간의 기억 능력은 제한적이기에 단순히 정보를 습득하고 분석하여 활용하는 것에서 그치는 것이 아니라 학습 과정이 더 다양해질 수 있다. 인간의 기억은 크게 단기기억과 장기기억으로 나눌

수 있는데, 단기 기억의 경우 시간이 지날수록 잊어버리는 일반적인 기억이다. 단기 기억은 시간이 흐르며 점점 잊거나 변형되기에 그 전 경험을 통해 후의 다른 상황에서 정확한 판단을 내리는 데 있어서 오류가 일어날 수 있다. 장기 기억의 경우 평생 유지가 가능할 수도 있는 기억이지만, 단기 기억을 장기 기억으로 변환시키기 위해선 많은 반복이 필요하다. 이 자극에 대한 분석을 계속 기억하기 위한 반복 과정에서 기존에 가지고 있던 기억에 대한 변화와 왜곡이 계속 일어나며, 이 변화는 개개인마다 다양한 학습 과정이 일어나게 하는 핵심적인 현상이다. 인간은 같은 정보를 받아들여더라도 원래 가지고 있던 기억의 차이와, 받아들인 정보를 해석하고 장기 기억으로 변환하는 과정에서 일어나는 왜곡과 변화의 차이로 다른 학습 결과를 만들어낸다.

IV. AI와 인간의 학습 과정의 차이

AI와 인간의 학습 과정의 차이는 크게 4가지로 볼 수 있다. 먼저 AI가 수집한 데이터에 대한 분석을 무조건 저장시켜 활용할 수 있다는 점과 달리 인간은 그 기억에 변화가 있다는 점이다. AI가 데이터를 분석한 결과값은 또 하나의 다른 데이터로서 컴퓨터의 하드웨어에 저장되어 필요할 때 다시 불러올 수 있다. 그러나 인간은 단기 기억의 경우 시간이 지날수록 잊어버리며 원래의 판단과 달라질 수 있으며, 장기 기억으로 변환하기 위해 반복을 하는 과정에서 계속 인지한 것에 대한 변화가 일어날 수 있다. 두 번째는 입력/받아들인 정보에 대한 해석의 차이이다. AI는 같은 시스템으로 이루어져 있을 경우 입력한 데이터에 대한 분석값이 항상 일정하게 나오고, 다른 시스템에 기반한 AI라도 그 차이는 거의 없는 편이다. 반면에 인간은 자극을 맥락에 따라 다르게 해석하고 받아들인다. 인간이 자극을 받아들이는 데 영향을 미치는 요소는 그 전까지 쌓여 왔던 기억, 즉 경험과 자극이 일어나는 상황이다. 모든 사람들은 살아온 삶이 다르기에 경험이 다르고, 그 자극이 일어나는 상황도 당연히 사람마다 다를 것이다. 동일 인물이 똑같은 자극을 받아들여더라도 그 상황에 따라 다르게 인지할 수 있기에 똑같은 데이터를 입력하면 똑같은 분석을 하는 AI와 다르게 인간은 다양한 해석이 나타난다. 세 번째 차이는 용량의 차이로, 정부 수용이 제한적인 인간과 비교했을 때 AI는 거의 무한한 양의 정보를 수용 가능하다. 때문에 무분별한 데이터 수집이 일어나는 AI는 학습 과정에서 주체성과 선택성이 없다고 볼 수 있는 반면, 인간은 정보를 수용할 수 있는 기억의 양도 부족하고, 받아들일 수 있는 정보의 시간적, 체력적 한계 또한 존재하기에 목적에 맞는 소량의 정보를 받아들이고 기억하는 과정에서 선택성과 학습의 주체성이 드러난다. 마지막으로 맥락, 가치관, 기존의 기억에 따라 자극 또는 정보가 옳고 그른지를 판단 가능한 인간과 달리 AI는 그대로 받아들인다는 점이다. 실제 'ChatGPT'의 가장 큰 문제점 중 하나로, AI가 정보의 참 거짓을 판단하지 못한 채 잘못된 정보를 올바른 정보처럼 인간에게 전달하는 예시가 있듯이, AI는 학습 과정에서 목적에 맞는 데이터를 수집할 때 그 데이터가 거짓이거나, 사실 목적에 부합하지 않는 데이터일지라도 여과 없이 수집하여 분석을 내린다. NovelAI가 일본 만화풍에 맞는 그림 데이터를 학습을 위해 수집할 때 일본 만화풍이라는 코드를 포함하고 있지만 코드만 그렇고 실제로는 일본 만화풍이 아닌 그림이더라도 AI는 일본 만화풍이라는 코드를 보고 목적에 맞는 데이터로 인식한다는 점이 그 예시가 될 수 있다.

V. 결론

AI가 그림을 그리기 위한 학습 과정에서 불법적 크롤링으로 대표되는 무단으로 데이터를 수집하여 사용하는 저작권 문제가 발생한다. 이는 이미 세계적으로도 'NovelAI'와 같은 여러 예시를 통해 대두되고 있음을 알 수 있는 문제이다. 다만 AI와 유사하게 인터넷에 돌아다니는 그림을 무단으로 따라 그리며 그림 실력을 키우는 인간의 학습 과정에는 저작권 문제가 왜 발생하지 않는지 AI와 인간의 학습 과정을 알아보았다. AI와 인간 모두 받아들인 정보를 분석, 판단하고, 이 정보에 대한 분석을 행동에 활용하는 것은 동일하다. 그러나 4장에서 알아보았던 몇 가지 차이들로 인해 왜 AI의 학습 과정에만 저작권 문제가 일어나는지를 알 수 있다.

AI는 입력되는 데이터가 맥락의 영향을 받지 않고 그대로 철저히 규칙에 따라 입력되고 판단되기 때문에 기존의 데이터를 그대로 저장하는 것과 차이가 없다고 볼 수 있다. 반면에 인간은 맥락에 따라 자극과 정보가 항상 다르게 해석되어 저장되고, 저장된 데이터의 변화가 없는 AI와 달리 단기 기억의 왜곡과 상실, 장기 기억을 위한 반복 과정에서의 변화를 통해 받아들인 정보 자체가 아닌, 자신의 것으로 변환된 정보를 가지고 있다는 것에서 차이가 나타난다. 때문에 대량의 데이터를 무분별하게 그대로 받아들이고 활용하는 AI는 저작물의 저작권을 침해한다고 볼 수 있지만, 인간의 경우 용량의 한계로 인한 소수의 선택적인 정보 습득과, 그 정보의 변화를 통해 정보 습득 과정에서 주체성이 드러나고, 또한 기존의 정보와는 다른 자신만의 정보를 만들어내기에 이 과정에서 저작권 문제에 대한 차이가 발생한다. 받아들인 정보를 창의성, 유추, 변화 없이 기존 그대로 수치화시키고, 그 수치들을 조합하여 그림을 만들어내는 AI는 주어진 데이터 외에 새로운 것을 생성하지 못하지만, 인간은 정보의 맥락에 따른 다른 판단과 분석, 그리고 변화 과정을 통해 새로운 데이터를 생성해낸다. 때문에 인간은 누군가의 작품을 보고 따라 그리며 연습했다 하더라도 자신만의 처리과정, 변화, 왜곡이 담긴 새로운 의미를 갖는 결과물을 만들어낸다. 결국 데이터를 바로 받아들여 저장하는 AI의 학습 과정과 맥락에 따라 다르게 인식하고, 또 왜곡하며, 양의 한계로 일어나는 정보의 선택과 반복을 통해 정보를 자신에게 맞도록 변화시키는 인간의 학습 과정에는 분명한 차이가 있기에 인간의 학습과는 달리 AI의 학습에서만 저작권 문제가 발생한다.

[참고문헌]

1. 학술논문

박현경(2020), 「인공지능 학습과정에서 저작물의 이용에 관한 소고」, 『스포츠엔터테인먼트와 법』, 23권 1호, 한국스포츠엔터테인먼트법학회, 131~134쪽.

차상욱(2017), 「인공지능(AI)과 지적재산권의 새로운 쟁점 - 저작권법을 중심으로-」, 『법조』, 66권 3호, 법조협회, 183~235쪽.

2. 기사 및 기타

김영대(2022. 11. 25.), 「그림·소설·영상까지… 예술 넘보는 인공지능」, 연합뉴스, http://www.yonhapmidas.com/article/221206113324_139276(2023. 05. 15).

한국저작권위원회(2023. 03. 15.), 「2023년 3월호(제343호) 이달의 핫픽 - 학습용 데이터와 AI 저작권 침해」, 한국저작권위원회 블로그, https://blog.naver.com/kcc_press/223044972630(2023. 05. 16.).

귀하가 제출한 본 서면은 교육 목적의 교재 제작과 수업 자료 등으로 사용될 수 있습니다.

이에 대해 사용 승낙을 하시겠습니까?

예

아니오

-우수과제 응모-